



TITLE:

# 死亡傾向の地域集積性に関する検定 (生命科学データの統計解析と数学モデル)

AUTHOR(S):

福富, 和夫; 永井, 正規

---

CITATION:

福富, 和夫 ...[et al]. 死亡傾向の地域集積性に関する検定 (生命科学データの統計解析と数学モデル). 数理解析研究所講究録 1980, 384: 77-91

ISSUE DATE:

1980-05

URL:

<http://hdl.handle.net/2433/104855>

RIGHT:

## 死亡傾向の地域集積性に関する検定

国立公衆衛生院 福富 和夫

永井 正規

### 1.はじめに

罹患ならびに死亡の地域傾向を観察することは、疾病要因の疫学研究の中でも、もっとも基本的な研究方法として、広く行われている。従来、主として府県レベルでなされてきた。この種の観察も、近年は市町村などの小地域情報が利用されるにおよび、一層重要な意義をもつものとなった。

研究のねらいと観察対象の規模により、疾病発生 の 地 域 的 観 察 を つ ぎ の よ う に 分 け る こ と が で き る 。 (1) 大 域 的 傾 向 の 観 察、たとえば、東日本と西日本、山内都と海沿いなどの比較。(2) 高率発生地区の集積性。(3) 孤立した高率発生地区の発見。

このうち、ここでは(2)の集積性の問題を取り上げる。集積性とは、高率発生地区が互に隣接する傾向をいう。このような場合、その地域一帯に疾病発生 の 要 因 の 偏 在 が 懸 念 さ れ る であらう。

集積性の研究には 2つの観点があり、1つは疾病の疫学的特性の研究（たとえば、がんは発生部位により、地域集積性のめられるものとそうでないものがある）、他は集積地域の確定である。いずれの場合も疫学地図を作ることにより、地域集積性はおおよそ視覚的に把握できるであろうが、偶然生じる集積も考えられ、検定の必要性が問題になる場合もある。

疾病の地域集積性に関する検定については、Geary (1954)<sup>1)</sup>、大野ら (1979a)<sup>2)</sup> (1979b)<sup>3)</sup>、永井ら (1979)<sup>4)</sup> の研究があるが、ここでは、これらを含む方法を示し、事例について相互を比較検討する。

なお、類似の問題に、時間・空間の集積性に関する検定の問題があり、Knox の方法<sup>5)</sup>、Mantel の方法<sup>6)</sup>などが知られているが、ここで述べる地域集積性の検定の一部は、Mantel の検定と数学的に同等のものになっている。

## 2. 地域集積性の尺度

観察対象地域は  $n$  個の小地域に分割されているものとする。 $p_i$  は小地域  $i$  における疾病指標（死亡率、罹患率など）、 $f_{ij}$  は小地域  $i$  と  $j$  の間の疾病発生傾向の類似度を表わす関数、 $d_{ij}$  は両地域の近さを表わす関数とする。ただし、 $f_{ij}$  は類似

しているほど、また、 $d_{ij}$  は近接するほど、大きな値をとるものとする。

地域集積性の尺度として、つぎのものを考える。

$$(1) \quad X = \sum_{i,j} d_{ij} f_{ij}.$$

ここで  $\sum_{i,j}$  は  $n$  小地域の全ての組合せについて和をとることを意味する。

いま、疾病指標  $p_i$  ( $i=1, \dots, n$ ) が与えられた条件の下で、帰無仮説「疾病指標は地域的にランダムに分布する」の検定を考える。このため、まず、仮説の下での  $E\{X\}$ ,  $V\{X\}$  を求めることにする。

$$E\{X\} = \sum_{i,j} d_{ij} E\{F_{ij}\}.$$

$E\{F_{ij}\}$  は  $p_i$  を各地域に等確率で割当てた場合の平均値 ( $i, j$ ) には依存しない。よって、添字を除き、 $E\{F\}$  と表わす。

$$E\{F\} = \sum_{i,j} f_{ij} / N, \quad \text{ただし} \quad N \equiv \binom{n}{2}.$$

つぎに、

$$(2) \quad V\{X\} = \sum_{i,j} d_{ij}^2 V\{F_{ij}\} + \sum_{i,j} \sum_{k,l} d_{ij} d_{kl} \text{Cov}\{F_{ij}, F_{kl}\}$$

ここで、 $E\{F\}$  の場合と同様の理由から、 $V\{F_{ij}\}$  の添字を除き、

$$V\{F\} = \sum_{i,j} f_{ij}^2 / N - \mu^2$$

と置く。ただし、 $\mu \equiv E\{F\}$ 。

(2) 式の第2項は、添字の組  $(i, j)$ ,  $(k, l)$  の間に共通なもの

が含まれる場合と、共通なものがない場合とに分けて扱うことになる。

前者には  $nP_3$  通りの組があり、共分散は

$$\text{Cov}\{F_{ij}, F_{ik}\} = \sum_{(i,j,k)} f_{ij} f_{ik} / n(n-1)(n-2) - \mu^2$$

となる。この式1項の分子は、つぎのように変形できる。

$$(3) \quad \sum_{(i,j,k)} f_{ij} f_{ik} = \sum_i f_{i\cdot}^2 - 2 \sum_{i,j} f_{ij}^2, \quad \text{ただし } f_{i\cdot} = \sum_{j \neq i} f_{ij}.$$

後者には  $\binom{n}{2} \times \binom{n-2}{2}$  通りの組があり、共分散は

$$\text{Cov}\{F_{ij}, F_{kl}\} = \sum_{i,j \neq k,l} f_{ij} f_{kl} / \binom{n}{2} \binom{n-2}{2} - \mu^2$$

となる。この式1項の分子は

$$(4) \quad \sum_{i,j \neq k,l} f_{ij} f_{kl} = \left( \sum_{i,j} f_{ij} \right)^2 - \sum_{i,j} f_{ij}^2 - \sum_{(i,j,k,l)} f_{ij} f_{kl}$$

と変形できる。また、(2)式1項2項の  $\sum_{i,j} \sum_{k,l} d_{ij} d_{kl}$  は、それぞれ対応する(3)、(4)の  $f \in d$  に置き換えて計算できる。

$$\sum_{(i,j,k,l)} d_{ij} d_{kl} = \sum_i d_{i\cdot}^2 - 2 \sum_{i,j} d_{ij}^2,$$

$$\sum_{i,j \neq k,l} d_{ij} d_{kl} = \left( \sum_{i,j} d_{ij} \right)^2 - \sum_{i,j} d_{ij}^2 - \sum_{(i,j,k,l)} d_{ij} d_{kl}.$$

集積性尺度としては、 $\Sigma/E\{X\}$  の形による表現が便利

な場合もある。集積性がみられるとき、この他は1より大となる。

検定統計量としては、

$$t = (\bar{x} - E\{x\}) / \sqrt{V\{x\}}$$

を用いる。nが大きく、かつ、 $E\{x\}$  もある程度大きいときは、帰無仮説の下で  $t$  が  $N(0,1)$  に従うとして検定できる。

### 3. 他の検定法との関係

#### (1) 大野の方法

これは  $d_{ij}$ ,  $f_{ij}$  をつぎのように定めた場合に対応している。

$$\begin{aligned} d_{ij} &= 1 && i \text{ 地域, } j \text{ 地域が隣接する場合} \\ &= 0 && \text{その他} \end{aligned}$$

$$\begin{aligned} f_{ij} &= 1 && p_i, p_j \text{ が同一階層区分に属する場合} \\ &= 0 && \text{その他} \end{aligned}$$

このとき、(1)の統計量は帰無仮説の下でパラメター  $\lambda = E\{x\}$  の Poisson 分布で近似できるとして検定する。この検定は数学的に Knox による時・空間集積性の検定と同等である。

#### (2) Gearyの方法

$d_{ij}$  の定義は上と同じで、 $f_{ij}$  はつぎのように定義してい

3.

$$f_{ij} = (p_i - p_j)^2$$

この場合は  $p_i$  と  $p_j$  の値が類似するほど、 $f_{ij}$  の値は小となる。  
検定には、

$$t = (1 - Z) / \sqrt{V\{Z\}}$$

を用いる。ただし、 $Z = \chi / E\{\chi\}$  ,  $\chi = \sum_{i,j} d_{ij} (p_i - p_j)^2$  .

また、 $V\{Z\}$  の近似式として、つぎのものを与えている。

$$V\{Z\} \doteq \frac{2}{n k_1^2} \{ k_2 + k_1 - k_1^2 \} ,$$

ただし、 $k_1 = \sum_{i=1}^n d_{i.} / n$  ,  $k_2 = \sum_{i=1}^n d_{i.}^2 / n$  ,  $d_{i.} = \sum_{j=1}^n d_{ij}$  .

上の  $V\{Z\}$  の近似式では、 $p_i$  の分布で Kurtosis がほぼ 0 とみなせることを仮定している。

### (3) 永井らの方法

$d_{ij}$  をつぎのように定義している。

$$d_{ij} = 1 / (\alpha D_{ij} + 1)$$

ここで、 $D_{ij}$  は  $i, j$  両域の中心間の距離である。地域の中  
心は市町村役場の所在地または地理的な中心をあらわしている。

$\alpha$  は任意に定める定数で、この値を変えることにより、近接  
度を変化させることができる。

$f_{ij}$  についても、

$$f_{ij} = 1 / (\alpha |p_i - p_j| + 1), \quad (\alpha \text{ は定数})$$

$$f_{ij} = \min(p_i, p_j) / \max(p_i, p_j)$$

など、種々の形を与えている。

#### 4. 3連結による集積性尺度

より密な集積性を表わすものとして、3の(1)のケースを下記のように変形した尺度も考えることができる。

小地域  $i, j, k$  が互に隣接したとき  $d_{ijk} = 1$ , その他は 0.  
小地域  $i$  が高率発生地区のとき  $p_i = 1$ , その他は 0 とし、 $f_{ijk} = p_i p_j p_k$  とする。集積性尺度は

$$X = \sum_{i,j,k} d_{ijk} f_{ijk}$$

とする。

以下、2連結の場合と同様に、帰無仮説の下で

$$E\{X\} = \sum_{i,j,k} d_{ijk} E\{F\}, \quad E\{F\} = \sum_{i,j,k} f_{ijk} / \binom{m}{3},$$

この場合、 $\sum_{i,j,k} f_{ijk} = \binom{m}{3}$  となる、ただし  $m$  は高率地区の個数。

$$V\{X\} = \sum_{i,j,k} d_{ijk}^2 V\{F\} + \sum d_{ijk} d_{i'j'k'} \text{Cov}\{F_{ijk}, F_{i'j'k'}\}$$

$$V\{F\} = \mu(1-\mu), \quad \text{ただし } \mu \equiv E\{F\}.$$



共分散は3つのケースに分けられる。

i)  $(i, j, k), (i', j', k')$  に等しいものが2対ある場合

$$\text{Cov}\{F_{ijk}, F_{ijk'}\} = mP_4 / nP_4 - \mu^2$$

$$R_1 \equiv \sum d_{ijk} d_{ijk'} = \sum_{ij} (d_{ij\cdot})^2 - 3 \sum_{ijk} d_{ijk}^2,$$

$$\text{ただし, } d_{ij\cdot} = \sum_{k \neq i, j} d_{ijk}.$$

ii)  $(i, j, k), (i', j', k')$  に等しいものが1対ある場合

$$\text{Cov}\{F_{ijk}, F_{i'j'k'}\} = mP_5 / nP_5 - \mu^2$$

$$R_2 \equiv \sum d_{ijk} d_{i'j'k'} = \sum_i d_{i\cdot\cdot}^2 - 2R_1 - 3 \sum_{ijk} d_{ijk}^2$$

$$\text{ただし, } d_{i\cdot\cdot} = \sum_{\substack{j < k \\ j, k \neq i}} d_{ijk}.$$

iii)  $(i, j, k), (i', j', k')$  に等しいものがない場合

$$\text{Cov}\{F_{ijk}, F_{i'j'k'}\} = mP_6 / nP_5 - \mu^2$$

$$R_3 \equiv \sum d_{ijk} d_{i'j'k'} = \left(\sum_{ijk} d_{ijk}\right)^2 - R_1 - R_2 - \sum_{ijk} d_{ijk}^2.$$

検定統計量は2連結の場合と同様に考えることができる。

## 5. 実例

上記の検定法を用い、肺がん、胃がん、自殺の3つの死因

について、秋田、岩手の両県の市町村に集積性がみられるかどうか、分析した結果を述べる。

資料は、胃がんについては、昭和44~52年、他の2死因については、昭和44~49年の人口動態統計による死亡数の合計および国勢調査(45年, 50年)人口である。死亡指標には、SMR (Standard Mortality Ratio)を採用した。これは、各小地域の総死亡数を、全国の年齢階級別死亡率に基づき期待死亡数で除し、100倍したものである。

死亡傾向の類似度の指標  $f_{ij}$  には、

$$(1) \quad f_{ij} = 1 \quad i, j \text{ 地域とも } SMR > 140 \text{ の場合} \\ = 0 \quad \text{その他}$$

$$(2) \quad f_{ij} = 1 \quad i, j \text{ 地域とも } SMR \text{ が } 100 \text{ より 有意に高い場合, ただし 有意水準は } 5\% \\ = 0 \quad \text{その他}$$

$$(3) \quad f_{ij} = 1 / (|p_i - p_j| + 1) \quad p_i \text{ は } SMR \text{ を表わす.}$$

など、3種を採用した。

近接基準  $d_{ij}$  は、いずれの場合も隣接のとき 1、その他は 0 という基準を用いている。

表1に検定結果を示す。また、図1はSMRで4階級に区分した疫学地図である。

表1.aより、 $f_{ij}$ の与え方が検定に大きく影響することが分るが、これはむしろ当然といえる。このうち、(1)および(2)と、(3)との開きがとくに目立つが、前者の $f_{ij}$ は死亡指標の値が両地域とも高い場合に、1となるのに対し、後者の $f_{ij}$ では死亡指標の値が類似した場合（必しも高い値とは限らない）に1に近い値をとる関数で、質的に異なるものである。

秋田の胃がんは、(1)および(3)で有意になっているが、図1の疫学地図をみても、高率地区の集積がはっきりとうかがえる。

表1.bは、Gearyの方法による結果を示すもので、全てのケースで有意になっている。これでは鋭敏すぎて、集積性を調べるための有用な尺度とは思えない。

表1.aの(1)および(2)列について $E\{X\}$ と $V\{X\}$ の値を比較すると、いずれの場合も $E\{X\} > V\{X\}$ の関係になっている。大野らの方法は、Poisson分布による近似を考え、 $V\{X\}$ の値を $E\{X\}$ で代用するものであるが、そのため検出力は若干低下することになる。

## 6. 考察

これまで、地域集積性について、その概念を明確にしないうまに論じてきたが、ここで簡単に触れておこう。疫学では、

ある疾患が、特定の集団（たとえば、地域人口集団、特定の家族など）に集中して多発する傾向がみられたとき、その疾患には集積性があるといふ。集団の種類によって、地域集積性、家族集積性などといふ。この場合、集積の原因には、遺伝、感染、生活環境など、様々なものが考えられるが、ここで問題とするのは、集積という現象の有無についてであり、原因の探究はつぎの研究段階に属するものである。

つぎに検定の問題であるが、発生頻度の低い疾患、もしくは急性疾患が特定の地域に限局して多発したような場合は、集積性は明白であろうから検定の必要性は起らないかもしれない。しかし、がんや循環器疾患などの慢性疾患が比較的に範囲において多発していると疑われる場合、既存の小地域情報を用いて高率地区の集積性が偶然によるものか否か、確かめることは疫学的に意味がある。とくに、人口の少ない町村の場合は死亡数、罹患数の変動が大きく、単独地域ごとの観察では正しい傾向性を把握できないことが多い。

永井と筆者は、地域の中心同の距離を基に、近接基準  $d_{ij}$  を定義したが（3節の(3)）、これはデータ処理が容易なばかりでなく、定数を変えることにより種々の度合の近接基準を与えられることができ便利である。

疾病指標の類似度の関数  $f_{ij}$  は、単に類似を表わすものと、

共に高率であることを表わすものとは区別できる。当然ながら後者の方が疫学地図による集積の視覚的印象に近い結果が得られる。

4節で与えた3連結による集積性尺度は、より密な集積性をねらったものであるが、計算時間が大幅に増大するため、このままの形では实际的ではない。何らかの近似式が必要であろう。

検定では、統計量の正規近似を用いているが、この近似が適切でない場合もある。たとえば、表1.2の肺がんでは  $f_{ij}$  が(2)のケースは  $E\{X_i\}$  の値が小さ過ぎる。このような場合は Poisson 近似による方がよからう。この統計量の高次の積率などのようになるかは、今後の検討課題である。

## 7. まとめ

疾病発生の小地域情報に基づく地域集積性の検定法を与え、実例について他法との比較を行なった。方法の有用性については、類似度の関数  $f_{ij}$  ならびに近接基準  $d_{ij}$  の選択など、さらに研究を要するが、疫学研究において十分意味のあるものと考えられる。

最後に、本研究について有益な助言を下された九州大学理学部工藤昭夫教授に、心から謝意を表します。

### 参 考 文 献

- 1) Geary, R.C. (1954), *Incorp Statist.*, 5, 115-145.
- 2) 大野良之他 (1979), *日本衛生学会雑誌*, 34 (3), 512-520.
- 3) Ohno, Y. et al (1979), *International Journal of Epidemiology*, 8 (3), 273-281.
- 4) 永井正規 他 (1979), *第50回日本衛生学会総会講演集*, 610, 402.
- 5) Knox, E.G. (1964), *Applied Statistics*, 13, 25-29.
- 6) Mantel, N. (1967), *Cancer Research*, 27 (2), 209-220.

表1.a 集積性検定の結果 —  $f_{ij}$  の比較

死因 $f_{ij}$		岩 手			秋 田		
		(1) SMR>140	(2) SMR有意	(3) $1/( p_i - p_j +1)$	(1) SMR>140	(2) SMR有意	(3) $1/( p_i - p_j +1)$
肺癌 (男)	E(x)	3.57	0.238	8.11	4.01	0.437	8.71
	V(x)	3.06	0.218	1.20	3.31	0.396	1.50
	$\chi$	4	1	8.43	7	1	12.0
	t	0.25	1.63	0.30	1.64	0.89	2.69**
胃癌 (男)	E(x)	0	0	12.9	13.9	0	12.2
	V(x)	0	0	1.91	9.91	0	1.97
	$\chi$	0	0	17.0	21	0	15.4
	t	—	—	2.94**	2.27*	—	2.23*
自殺 (男)	E(x)	34.5	7.22	6.34	40.9	4.81	6.81
	V(x)	21.4	5.91	1.15	21.8	3.91	1.03
	$\chi$	52	19	8.16	45	6	7.30
	t	3.78**	4.84**	1.70	0.88	0.60	0.48

\*  $P < 0.05$     \*\*  $P < 0.01$ 

表1.b Gearyの方法による検定結果

死因		岩 手	秋 田
肺癌 (男)	$Z$	0.496	0.498
	t	4.71**	5.48**
胃癌 (男)	$Z$	0.401	0.615
	t	5.60**	4.03**
自殺 (男)	$Z$	0.251	0.505
	t	7.01**	5.19**
V(Z)		0.0114	0.00911

\*\*  $P < 0.01$

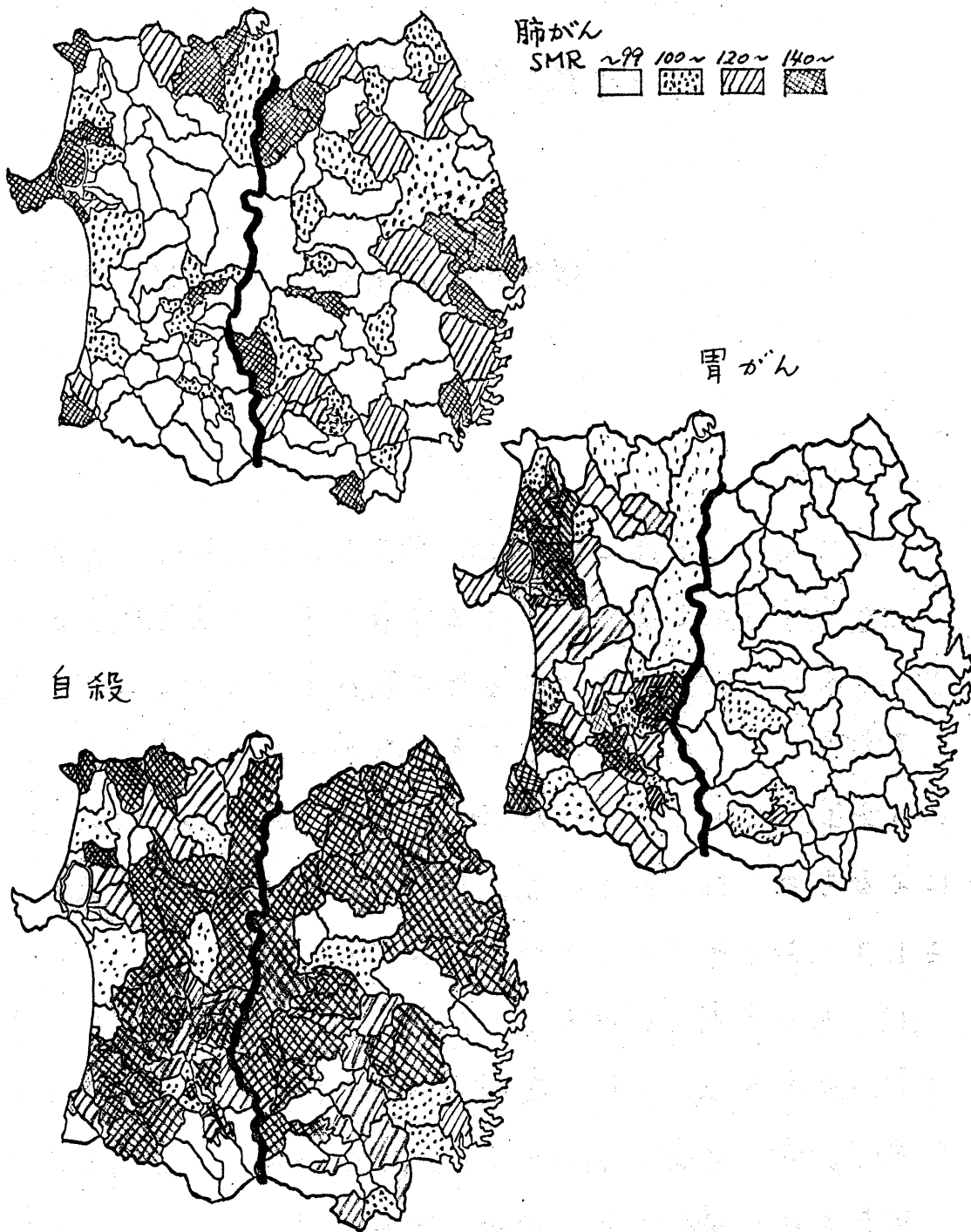


図1 肺がん・胃がん・自殺の疫学地図